Completing a Data Management Plan

Postgraduate Research Students

Introduction

This document is intended to help you complete a data management plan, and gives you some extra things to consider that aren't mentioned in the template. You might not know the answers to everything — if there's something you're not sure about, make a note on the plan to find out. You may wish to discuss some or all of it with your supervisor.

Type as much (or as little) as you feel you need to into each box: it will expand to accommodate what you write. The *text in grey* on the template gives examples of possible answers — use or replace it as needed.

Data Management Plans, Data Sharing Statements and other similarly-named documents are now required by most research funders. Increasing numbers of universities require them as well. Those required by funders often go into more detail about staff and resources than this template for PGRs, but this template has been specifically designed to help you plan the aspects of data management relevant to a doctoral research project.

Completing each section

Overview

This section is for administrative purposes, to make it clear which project the plan is about. "Project context" need only be two or three sentences summarising your project's aims.

Defining your data

Almost all research builds on sources of information to develop and justify conclusions, whether this information is newly created or gathered from existing sources; this is your research data. This section helps you to think about what 'data' means for your research.

The answers you enter in this section will help you identify what resources, such as storage, you will need to manage your data. They will also help the University to understand and plan for growth in demand

Where does your data come from?

Is it gathered from experiments? From literature? From existing data on the web? Or from somewhere else?
What instruments do you use? How about observations or photos?

How often do you get new data?

Continuously over a long period or from separate one-off events such as experiments or interviews? How many experiments per week? How will this change over time?

How much data do you generate?

Try to give this in kB/MB/GB. Start with how much you have so far, and try to estimate how this will grow for the rest of the project, based on your answer to the previous question.

If you keep data in a non-digital format, such as a lab notebook, consider how many notebooks you might need.

Tip: You can find out the size of an existing file or folder by right-clicking on it in Windows Explorer or Mac Finder and selecting Properties… (Windows) or Get info… (Mac).

What data formats do you use?

What software is required to access and analyse the data? Is it free/open, and if not are alternatives available? How would you access your data if the university no longer had a license for the software you currently use?

What type of data does each format hold?

Looking after your data

The answers you enter in this section will help you identify how to keep your data safe, and will also make it easier for other researchers to make use of your data after the end of your project, if appropriate.

What different versions of each data file do you create?

In this context “versions” of a file doesn’t simply refer to multiple copies (such as you might make when backing up), but updated copies when the contents of a file are changed.

Do you update or add data to existing files, or do you create new files when you add new data? How do you indicate this in the filename?

Do you create additional files during analysis? If so, how do filenames from different stages of your analysis relate to each other?

What additional information is required to understand each data file?

What would you need to know to reproduce the data? If someone else in your lab or a reader of your papers wanted to replicate your analysis, what would they need to know? If you have used abbreviations or codes in your data, how will others know what they mean?

This type of detail is particularly important to record because it is often glossed over in published outputs, where the general method and conclusions are more important than the fine detail.

Once you’ve decided what information should be recorded, you should go ahead and record it in a “readme” file (or similar) that you store with your data. You could think about setting up a template to make this quicker for new data.

Where do you store your data?

If you have more than one copy of your data (say on a laptop and desktop computer) you should decide, early on, which is the primary copy, as this

avoids confusion later. It's surprisingly easy to lose data because you can't remember where you put it.

How do you structure and name your folders and files?

Planning this is another good technique to save time (and stress) later. You might wish to organise your data according to the date it was collected, or according to a key aspect of the conditions it was collected under.

How is your data backed up?

An ideal backup strategy should follow the 3-2-1 rule: at least 3 copies, on at least 2 different media (e.g. disk and tape), with at least 1 kept offsite. By far the easiest way to do this is to make the university's research storage (X: or H: drives) your primary storage area and let BUCS do the hard work for you. Don't forget that paper copies of data, such as notebooks, can be scanned directly to your University home drive.

How will you test whether you can restore from your backups?

Sometimes backups don't work as you expect, so you should check regularly that everything's correct. If you use the X: or H: drives, this could be as simple as checking that your files are up to date and will still open correctly.

Sharing your data

This section will help you understand how and when you should share your data. This is particularly important for publicly-funded research, as funders are increasingly expecting data to be published itself, in addition to traditional journal publications. Publishing data also helps to support the integrity of the academic record, because analyses described in papers can easily be reproduced.

Who owns the data you generate?

The university legally owns most data generated by students it funds, although your "scholarly output" such as presentations and articles remain yours. If you're not sure, check your studentship agreement or ask your supervisor.

Who else has a right to see or use this data?

Collaborators? Group members? Just your supervisor?

Who else should reasonably have access?

Readers of your published work? The general public?

Remember: many funders hold that publicly-funded data is a public good and should be treated as such.

What should/shouldn't be shared and why?

Sometimes there are good reasons for keeping even publicly-funded research data private, such as patient confidentiality or commercial sensitivity. If your original data can't be shared, consider whether a redacted or anonymised version might be shareable. A defined embargo period would allow you time to commercialise your research before it is made publicly available.

Archiving your data

This section will help you plan how you will ensure your research data is preserved at the end of your project. It will also help the university to plan for the ongoing preservation of our research data as a whole.

What should be archived beyond the end of your project?

Must everything be kept, or just what you used for your thesis? What secondary use might there be for your data, either on its own or combined with additional data? What might others need/want?

If your data is very large but easy to reproduce accurately (e.g. simulation results), consider whether recreating it or storing it for a long period would be the cheaper option. What *would* you need to keep so that you could guarantee the reproduction was accurate?

For how long should it be stored?

EPSRC guidelines state that data underlying published articles must be kept for 10 years from the date of last access, and other funders have similar policies.

Where will the archive be stored?

Many subject areas and disciplines have specialised archives to encourage the preservation and reuse of particular types of data. Your supervisor and others in your lab, group or department should be able to tell you if these exist. You may well have used such an archive to look up data for your own research.

When will files be moved into the archive?

As you complete the analysis of each file? When you publish a paper? When you submit your thesis?

Who is responsible for moving data to the archive and maintaining it?

You or your supervisor will probably be responsible for handing over data, but once there, the archive is likely to have procedures in place for maintaining it.

Who should have access and under what conditions?

Consider your answers to the previous section on data sharing. If your data can't be released immediately, could it be released at a later date, after a reasonable embargo period? If your data contains information that must be kept confidential, could access be granted to other researchers if they agree to certain conditions and, if so, what conditions are these?

Executing your plan

Who is responsible for making sure this plan is followed?

You may wish to discuss and agree this with your supervisor.

How often will this plan be reviewed and updated?

You may wish to discuss and agree this with your supervisor.

What actions have you identified from the rest of this plan?

List them here with timescales for their completion.

What further information do you need to carry out these actions?

Where can you find this information?

Who might you be able to ask?

Other sources of information

- University of Bath research data website: http://www.bath.ac.uk/research/data
- Digital Curation Centre resources: http://www.dcc.ac.uk/resources